

Attorney Docket No.: RADSA 21.075 (101120-00058)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Inventor(s): Philip DERBEKO  
Confirmation No.: 5259  
Serial No.: 10/813,757  
Filed: March 31, 2004  
Title: A METHOD FOR ACCELERATING THE EXECUTION...  
Examiner: Kaushikkumasr M. Patel  
Group Art Unit: 2188

August 8, 2007

Commissioner for Patents  
P.O. Box 1450  
Alexandria, VA 22313-1450

DECLARATION OF PRIOR INVENTION UNDER 37. C.F.R. § 1.131

SIR:

I, Gadi Erlich, hereby declare as follows:

1. I am the Chief Technology Officer and General Manager of Sanrad, Ltd., the assignee of the above-identified application. I am authorized to act on behalf of Sanrad, Ltd., in matters relating to the intellectual property of Sanrad, Ltd.

2. At the time of filing of the above-identified application I held the position of Vice-president of Research and Development at Sanrad, Ltd. In my position of Vice-president of Research and Development, I was tasked, among other things, with determining which developments were to be the subject of patent applications.
3. Phillip Derbeko, the inventor of the subject matter of the above-identified application, was previously, but is no longer, an employee of Sanrad, Ltd.
4. The above-identified application was filed with the United States Patent and Trademark Office on March 31, 2004.
5. On May 9, 2003, Mr. Derbeko provided me with a presentation (Attached as Exhibit A) entitled "Journaling" detailing the invention. This presentation served as an invention disclosure statement.
6. After review of the presentation, I directed the presentation to be forwarded to counsel for Sanrad, Ltd., for preparation of a patent application. The agents of Sanrad, Ltd., acted diligently from prior to May 9, 2003 until to filing of U.S. Application Serial No, 10/813,757 on March 31, 2004 to develop the invention and process and prepare the application.
7. I further declare that all statements made herein of my own knowledge are true and all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are

punishable by fine or imprisonment, or both, under section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the above-referenced application or any patent issuing thereon.

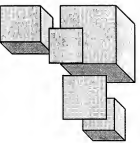
Respectfully submitted,

Date 8/8/07

  
\_\_\_\_\_  
Gadi Erlich

Chief Technology Officer and General Manager  
Sanrad, Ltd.

**EXHIBIT A**



Intelligent Storage Networks

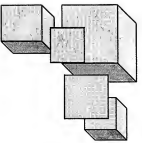
# Journaling

By

Philip Derbeko

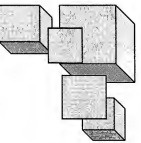
IF YOU'RE A SYSTEMS ADMINISTRATOR, YOU'VE GOT TO READ THIS BOOK. IT'S THE ONLY ONE YOU NEED.





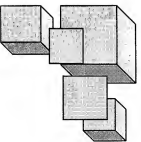
## Three uses of Journaling:

1. Remote Mirror (asynchronous)
2. Fast Mirror Sync
3. Mirror Data Consistency

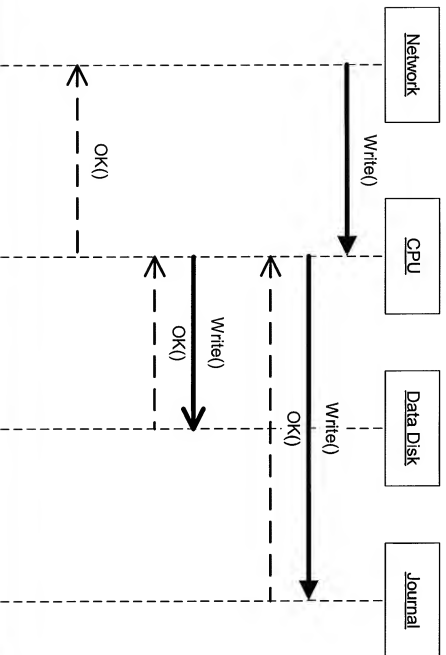


## **Remote Mirror (asynchronous)**

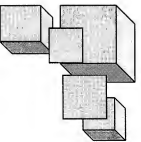
- The journal should be unlimited in a number of changes in other words no change should be lost even in expense of redundant synchronizations.
- Local writes should be as fast as possible.



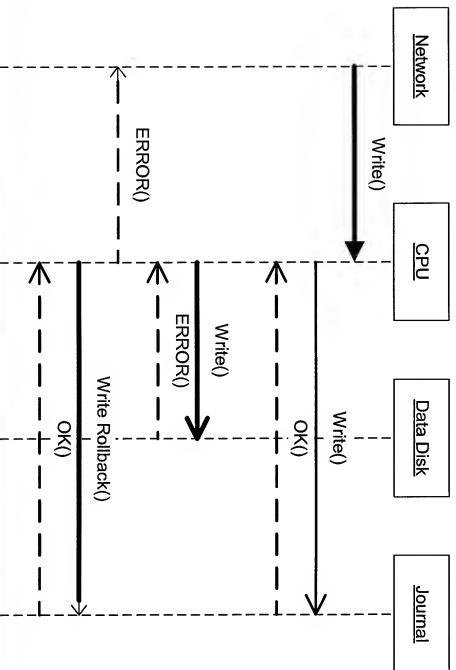
## Remote Mirror (asynchronous)

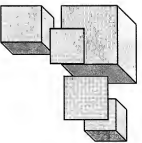






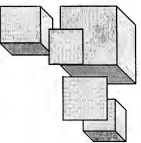
## Remote Mirror (asynchronous)



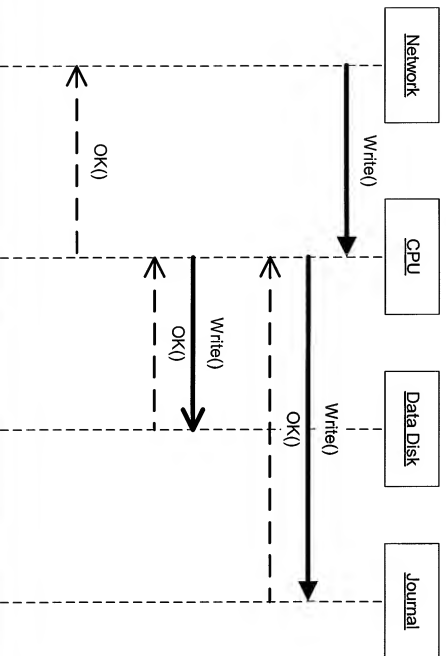


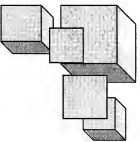
## Fast Mirror Sync

- The journal should be unlimited in a number of changes in other words no change should be lost even in expense of redundant synchronizations.
- Local writes should be as fast as possible.
- Can be used only to synchronize with the same disk.



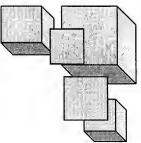
## Fast Mirror Sync



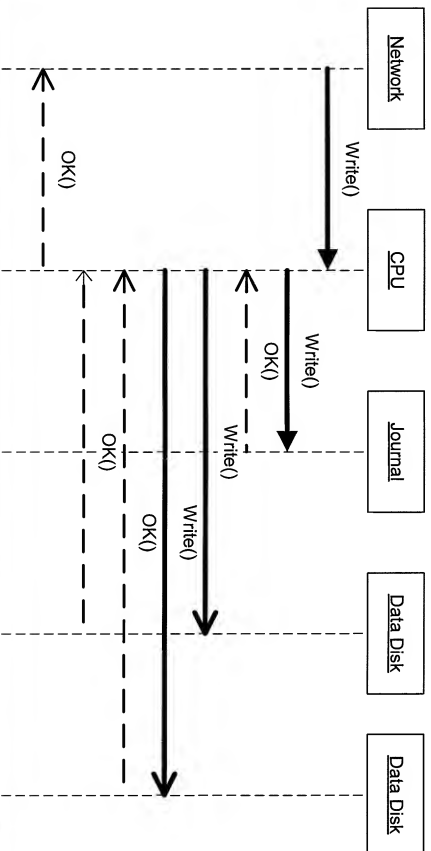


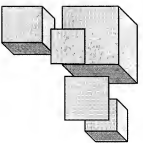
## Mirror Data Consistency

- The goal is to ensure that all mirror branches contain the same data all the time. The problem occurs when some of the commands that did not finish on all the branches and the VSwitch rebooted.
- The VSwitch should ensure that all mirror branches have the same data. The data can be chosen arbitrarily from any branch, of course, if we can do better so should we.
- The added value is unclear.

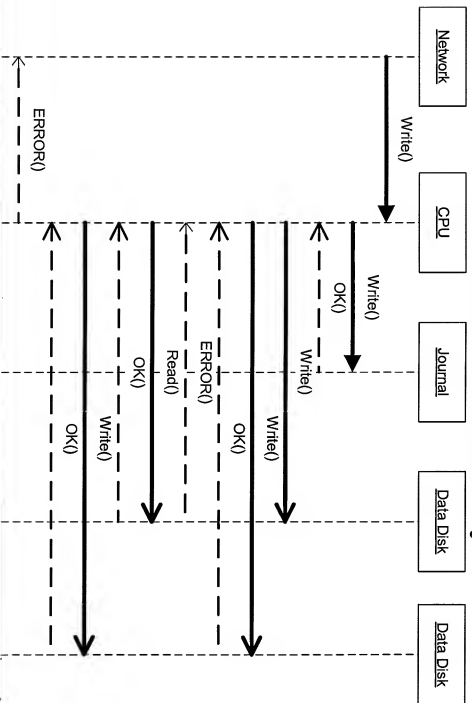


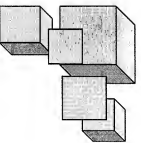
## Mirror Data Consistency





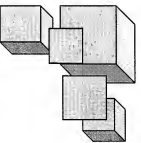
## Mirror Data Consistency





## Mirror Data Consistency

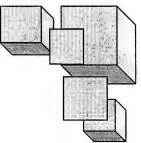
- Performance penalty: ALL writes latency is doubled, error writes latency is multiplied by 4 (at least).
- Complex implementation.



# Issues

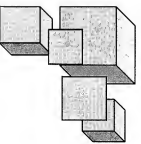
- What is written in the journal records ?
- Logical commands (only writes) to save space and to enable playback even from a volume even in cases when the hierarchy changed (add/break a mirror, volume resize).
- When is the journal record deleted ?
- After the synchronization completed; the synchronization might be between local disks or remote backup.
- States ?
- There is no need to keep the states in the journal. All we are interested is whether the change have been made (or attempted) or not.





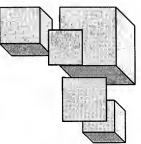
## Space Consideration

- What happens when the journal space runs out ?
- An SNMP trap is sent to the management station at about 80%.
- If the journal reaches 90% of a disk space ?
- The journal is “compressed” by merging a close by changes over the non-changed gaps. For example: given block changes 10-20 and 30-40, they can be merged into one single record : 10-40. The data integrity of a journal is not affected, though more data is sent during the synchronization.
- In the future ?
- To decrease the risk of the previous item, the journal data should be compressed (for example LZ) from the start.



## Where to keep the Journal?

NVRAM	Disk
Fast access	Slow access
Local to each VSwitch (should be synchronized in cluster)	Shared between all the VSwitches in cluster and even by boxes outside the cluster.
New hardware platform	Does not depend on hardware
Limited in size	Unlimited in size, the volume can be resized as it grows.
Write to journal always succeeds	Write to journal can fail.



## What kind of disk storage?

Storage Pool	Separate Volume
Transparent to user	User has to define the volume
Can grow/shrink dynamically	Requires manual interruption from administrator to grow/shrink
Demands more synchronization between Multi SDC	No synchronization is needed, all the data can be read from journal
Development time	Less development time
Less control from the administrator side	The administrator fully controls the size and location of a journal